Supplementary Materials: Cross-Granularity Online Optimization with Masked Compensated Information for Learned Image Compression

Haowei Kuang[1]   Wenhan Yang[2]   Zongming Guo[1]   Jiaying Liu[1]
[1]Wangxuan Institute of Computer Technology, Peking University, Beijing, China
[2]Pengcheng Laboratory, Shenzhen, China
kuanghw@stu.pku.edu.cn, yangwh@pcl.ac.cn, {guozongming, liujiaying}@pku.edu.cn

1. Model Details

In this section, we provide the hyper-parameters of our network. The hyper-parameters of the transform model and entropy model are same as [9], with channel numbers C of TCM blocks as 128. The hyper-parameters of our proposed MSC Encoder and MSC Decoder are shown in Table. 1, where nb denotes there is no bias in the convolution. The hyper-parameters of DDT Encoder and DDT Decoder are shown in Table. 2, which are based on [11] but with residual connection in DDT Decoder, where s2 denotes stride=2 in the convolution.

Table 1. Hyper-parameters of MSC Encoder and MSC Decoder.

MSC Encoder	MSC Decoder
Conv: 1×1 $320 \rightarrow 256$	Conv: 1×1 $320 \rightarrow 256$ nb
GeLU	GeLU
Conv: 1×1 $256 \rightarrow 512$	Conv: 1×1 $256 \rightarrow 512$ nb
GeLU	GeLU
Conv: 1×1 $512 \rightarrow 256$	Conv: 1×1 $512 \rightarrow 256$ nb
GeLU	GeLU
Conv: 1×1 $256 \rightarrow 320$	Conv: 1×1 $256 \rightarrow 320$ nb

Table 2. Hyper-parameters of DDT Encoder and DDT Decoder.

DDT Encoder	DDT Decoder
Conv: 3×3 $3 \rightarrow 32$ s2	FC: $9 \rightarrow 128$
ReLU	GeLU
Conv: 3×3 $32 \rightarrow 64$ s2	FC: $128 \rightarrow 256$
ReLU	GeLU
GlobalAvgPooling	FC: $256 \rightarrow 3 * 3 * 1 * 1$
Concat	
Conv: 1×1 $99 \rightarrow 9$	

2. Training Details.

In this section, we provide more specific training details. We use DIV2K image dataset [1] as our training dataset, which contains 800 high-quality natural images with an average 2K resolution. To enhance resolution adaptability, we implement data augmentation through bicubic downsampling on the images to half their original resolution. During all stages of training, the training patches are extracted through randomized cropping of 256×256 pixel regions from images. We use the Adam optimizer [6] in each phase of the training, with the initial learning rate set to 1×10^{-4}. For the four stages of training (DDT module training, MSC Encoder/Decoder training, codebook warmup training, codebook training), 500, 100, 200, 1000 epochs are trained, respectively.

3. Implementation of the Compared Methods

The code links for all the methods compared are listed in Table 4. For methods with published original RD data in their paper, we directly use these data for comparison. For the methods that expose the pre-trained model, we use their official released pre-trained models to perform inference on the test set to get comparative results. We use VTM-12.1 which is the official reference software to achieve VVC. And for BPG, we utilize BPG v0.9.8 with the quantizer parameters. Thanks to the authors for sharing their codes and pre-trained models, which are very helpful for our research.

4. Performance on Kodak dataset

We further show the performance on the dataset of Kodak [7] in Fig. 1 and Table 3. The performance is similar to the performance on CLIC professional dataset [10] which has been shown in the main paper.

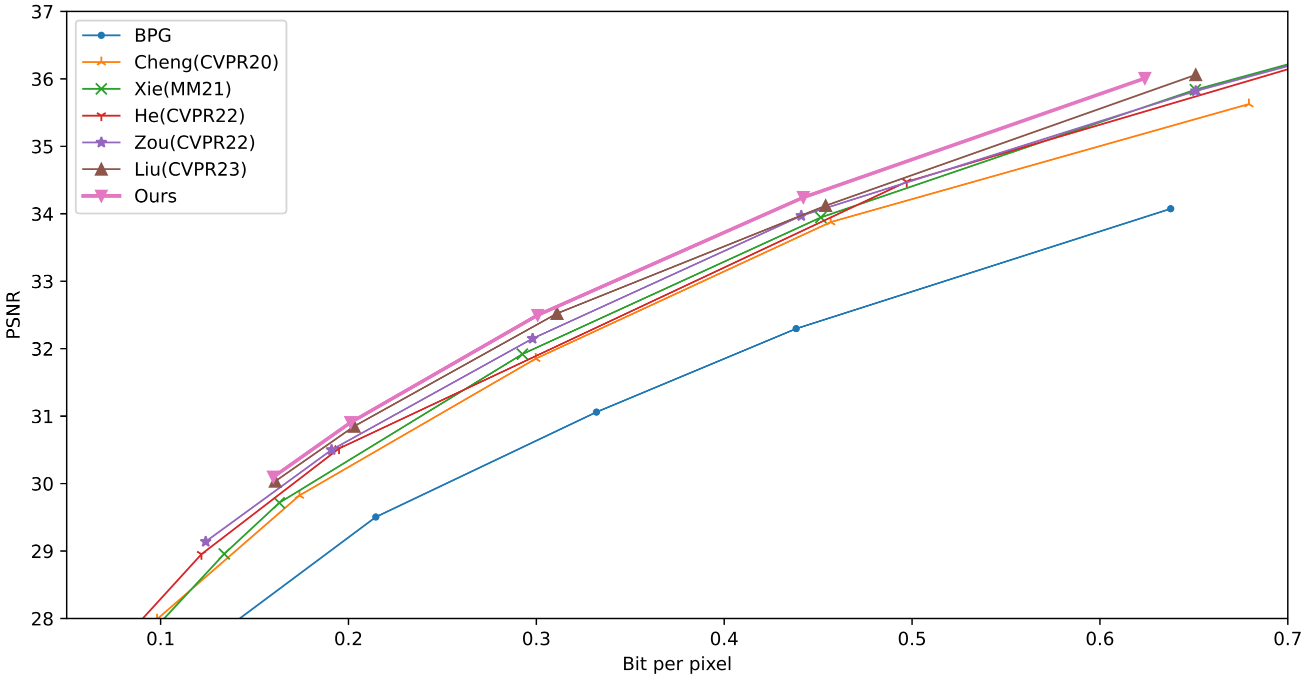


Figure 1. RD-Curve on Kodak dataset. [*Zoom in for best view*]

Table 3. BD-rate results on Kodak dataset [7]. We set BPG [2] as the anchor in the calculation. The best results are shown in **bold**.

Method	BD-Rate
Cheng (CVPR-20) [4]	-25.76%
Xie (ACMMM-21) [12]	-27.74%
He (CVPR-22) [5]	-31.58%
Zou (CVPR-22) [13]	-36.84%
Liu (CVPR-23) [9]	-33.23%
Ours	**-35.21%**

5. More Visual Results

More visual results of our methods compared with other methods are shown in Figs. 2 and 3.

Table 4. Code or original data links of the compared methods.

Method	Code Link
BPG [2]	https://bellard.org/bpg/
VVC [3]	https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/releases/VTM-12.1
Cheng(CVPR20) [4]	https://github.com/ZhengxueCheng/Learned-Image-Compression-with-GMM-and-Attention
Xie(ACMMM21) [12]	https://github.com/xyq7/InvCompress
He(CVPR22) [5]	https://github.com/VincentChandelier/ELiC-ReImplemetation
Zou(CVPR22) [13]	https://github.com/Googolxx/STF
Liu(CVPR23) [9]	https://github.com/jmliu206/LIC_TCM
Li(ICLR24) [8]	https://github.com/qingshi9974/ICLR2024-FTIC

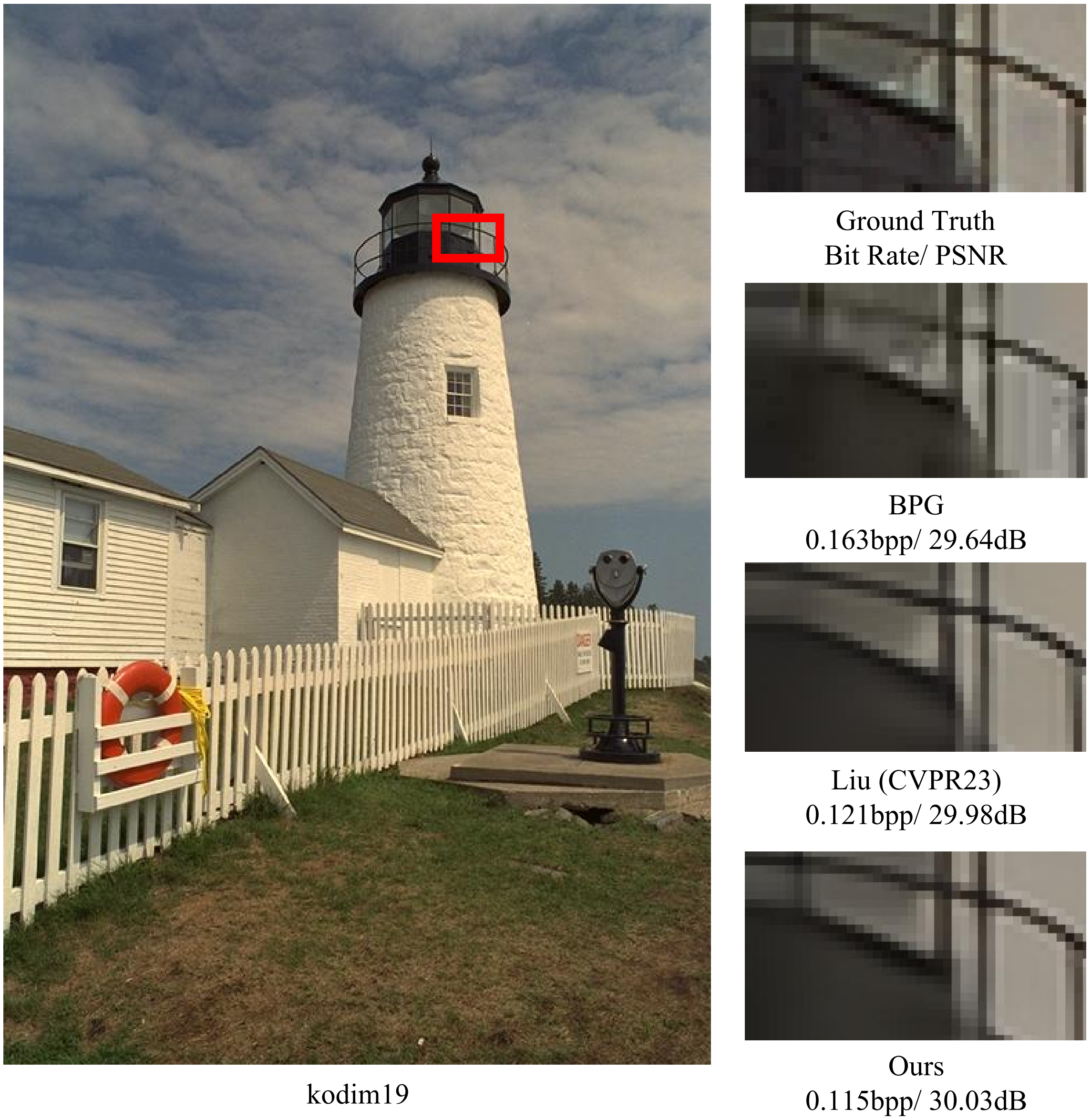
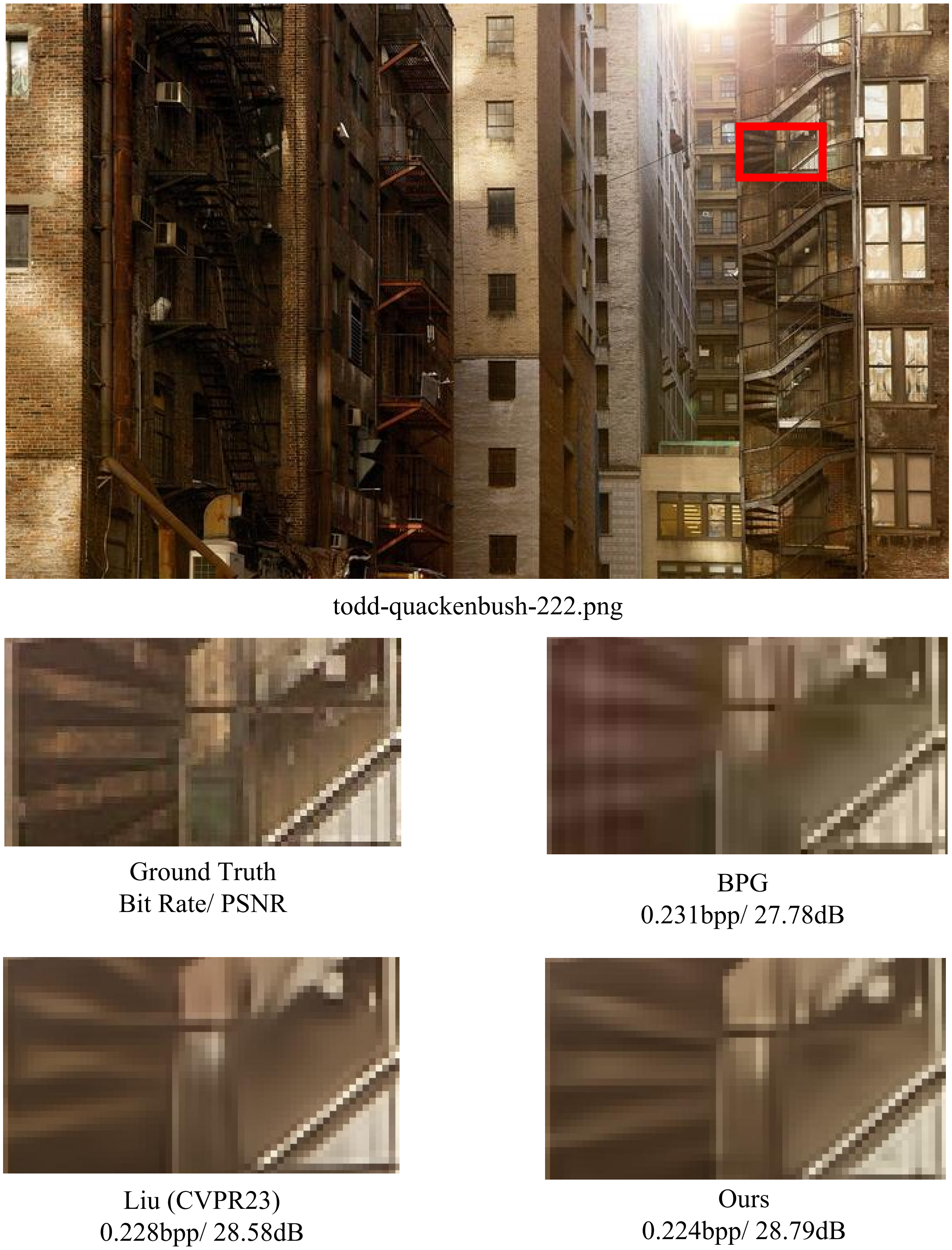


Figure 2. Subjective results on kodim19 from Kodak [7].



todd-quackenbush-222.png

Ground Truth
Bit Rate/ PSNR

BPG
0.231bpp/ 27.78dB

Liu (CVPR23)
0.228bpp/ 28.58dB

Ours
0.224bpp/ 28.79dB

Figure 3. Subjective results on todd-quackenbush-222 from CLIC dataset [10].

References

[1] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2017. 1

[2] Fabrice Bellard. BPG image format. http://bellard.org/bpg/, 2017. 2, 3

[3] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J Sullivan, and Jens-Rainer Ohm. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Trans. Circuit Syst. Video Technol.*, 31(10): 3736–3764, 2021. 3

[4] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2020. 2, 3

[5] Dailan He, Ziming Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. ELIC: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022. 2, 3

[6] Diederik Pieter Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, 2014. 1

[7] Eastman Kodak. Kodak lossless true color image suite. https://r0k.us/graphics/kodak/, 2013. 1, 2, 3

[8] Han Li, Shaohui Li, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong. Frequency-aware transformer for learned image compression. In *Int. Conf. Learn. Represent.*, 2024. 3

[9] Jinming Liu, Heming Sun, and Jiro Katto. Learned image compression with mixed transformer-CNN architectures. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2023. 1, 2, 3

[10] George Toderici, Wenzhe Shi, Radu Timofte, Lucas Theis, Johannes Balle, Eirikur Agustsson, Nick Johnston, and Fabian Mentzer. Workshop and challenge on learned image compression. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog. Worksh.*, 2020. 1, 4

[11] Dezhao Wang, Wenhan Yang, Yueyu Hu, and Jiaying Liu. Neural data-dependent transform for learned image compression. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022. 1

[12] Yueqi Xie, Ka Leong Cheng, and Qifeng Chen. Enhanced invertible encoding for learned image compression. In *ACM Int. Conf. Multimedia*, 2021. 2, 3

[13] Renjie Zou, Chunfeng Song, and Zhaoxiang Zhang. The devil is in the details: Window-based attention for image compression. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022. 2, 3